AMENDMENTS TO THE SPECIFICATION

Please amend the paragraph beginning on line 8 of page 16 as follows:

Techniques for determining nucleic acid and amino acid "sequence identity" also are known in the art. Typically, such techniques include determining the nucleotide sequence of the mRNA for a gene and/or determining the amino acid sequence encoded thereby, and comparing these sequences to a second nucleotide or amino acid sequence. In general, "identity" refers to an exact nucleotide-to-nucleotide or amino acid-to-amino acid correspondence of two polynucleotides or polypeptide sequences, respectively. Two or more sequences (polynucleotide or amino acid) can be compared by determining their "percent identity." The percent identity of two sequences, whether nucleic acid or amino acid sequences, is the number of exact matches between two aligned sequences divided by the length of the shorter sequences and multiplied by 100. An approximate alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman, Advances in Applied Mathematics 2:482-489 (1981). This algorithm can be applied to amino acid sequences by using the scoring matrix developed by Dayhoff, Atlas of Protein Sequences and Structure, M.O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA, and normalized by Gribskov, Nucl. Acids Res. 14(6):6745-6763 (1986). An exemplary implementation of this algorithm to determine percent identity of a sequence is provided by the Genetics Computer Group (Madison, WI) in the "BestFit" utility application. The default parameters for this method are described in the Wisconsin Sequence Analysis Package ProGram Manual, Version 8 (1995) (available from Genetics Computer Group, Madison, WI). A preferred method of establishing percent identity in the context of the present invention is to use the MPSRCH package of proGrams copyrighted by the University of Edinburgh, developed by John F. Collins and Shane S. Sturrok, and distributed by IntelliGenetics, Inc. (Mountain View, CA). From this suite of packages the Smith-Waterman algorithm can be employed where default parameters are used for the scoring table (for example, gap open penalty of 12, gap extension penalty of one, and a gap of six). From the data generated the "Match" value reflects "sequence identity." Other suitable proGrams for calculating the percent identity or similarity between sequences is generally known in the art, for example, another alignment proGram is BLAST, used with default parameters. For example, BLASTN and BLASTP can be used using the following default parameters: genetic code = standard; filter = none; strand = both; cutoff = 60; expect = 10; Matrix = BLOSUM62; Descriptions = 50 sequences; sort by = HIGH SCORE; Databases = non-redundant, GenBank + EMBL + DDBJ + PDB + GenBank CDS translations + Swiss protein + Spupdate + PIR. Details

of these proGrams can be found on the internet. ~~at the following internet address: http://www.ncbi.nlm.gov/cgi-bin/BLAST~~.

Please amend the paragraph beginning on line 2 of page 36 as follows:

The transposon cassettes are typically cloned into shuttle vectors for ease of manipulation and isolation of large quantities of vector DNA. A number of such shuttle vectors are commonly available, e.g., pAUL-A (Chakraborty, et al. (1992) *J. Bacteriol.* **174:**568-574) (a schematic diagram of the vector is presented in Figure 1); pE194 (Sozhamannan, s., et al. (1990) *J. Bacteriol.***172:** 4543-4548; see the ATCC website on the internet ~~http://phage.atcc.org/vectors/gifs/68359.gif~~ for a map of this vector; see the ATCC website on the internet ~~ftp://ftp.atcc.org/pub/vector_seqs/pE194.html~~ for the full sequence); pMK4 (Sullivan, M., *et al.,* (1984) *Gene* **29**:21-26), pDL289 (Buckley, N., *et al.,* (1995) *J. Bacteriol* **177**:5028-5034), pSK+ BLUESCRIPT(Clontech, Palo Alto, CA; Stratagene, La Jolla, CA); and the pSUM series mycobacteria shuttle vector (Ainsa, J.A., *et al.*, (1996) *Gene* **176**:23-26). In preferred embodiments, the shuttle vectors preferably include the following features: (1) a Gram-positive origin of replication, and/or a Gram-negative origin of replication, and/or an origin or replication functional in both types of organisms; (2) polylinkers; and (3) a polynucleotide encoding a selectable marker (*e.g.*, ampicillin, chloramphenicol, erythromycin, and others as discussed *supra*) which allows selection in the host cells. Most preferably, the shuttle vectors will further include (4) transcription termination sequences flanking one or both sides of the transposon cassette. Such transcription termination sequences are used to prevent transcriptional read-through into the coding sequences of the transposon cassette.